

Keywords extraction

Main plan: extract keywords from the text and visualize the keywords using various methods. Our input is in form of string, and the output can be a string, a text file or a plot, depending on the function we choose to use.

Accomplished work:

We compare the most popular keyword extractors (including 'tf-idf', 'kp-miner', 'rake', 'yake', 'text-rank', 'single-rank', 'topic-rank', 'multipartite-rank', 'keybert', 'count-vectorizer', 'openai').

```
res = ke.extract_keywords(text, num_keywords=10,  
method='keybert', save_txt=True)
```

allowed us to choose the method we want and whether we want to save the result as a text file.

For visualization, we have:

Basic word cloud;

n_grams_wordcloud which allow us to select the range of n grams;

get_relation_plot which analyze the subject, object and the relation in a sentence and output a relationship plot;

shell_diagram which connects all related entities. (Line segments are not weighted);

circosplot which connects all related entities. (Line segments are weighted);

packed bubble chart based on the frequencies or the keybert score of the top words;

multiple bar charts and combining them to save space;

treemap with the top 100 words of the selected topic;

Circle packing which create a bubble plot with no overlapping area.